To find least squares solution for $Ax = b$
(brute force method)

1) Find a basis for $W = \text{Col } A$: a basis consists of the pivot columns of A

2) Apply the Gram-Schmidt Process (*if necessary*) to this basis to get an orthogonal basis $\{\boldsymbol{v_1}, ..., \boldsymbol{v_i}, ..., \boldsymbol{v_p}\}$ for $W = \text{Col } A$

3) Use Orthogonal Decomposition Theorem to get

$$\widehat{\boldsymbol{b}} = \frac{\boldsymbol{b}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} + ... + \frac{\boldsymbol{b}\cdot\boldsymbol{v_i}}{\boldsymbol{v_i}\cdot\boldsymbol{v_i}}\boldsymbol{v_i} + ... + \frac{\boldsymbol{b}\cdot\boldsymbol{v_p}}{\boldsymbol{v_n}\cdot\boldsymbol{v_p}}\boldsymbol{v_p}$$

4) Solve $A\boldsymbol{x} = \widehat{\boldsymbol{b}}$

(*This is not so hard if A has orthogonal columns at the beginning: in that case, steps 1) and 2) are unnecessary – see first example*)

Example Find the least squares solution of

$$\begin{matrix} A \\ \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} \end{matrix} \boldsymbol{x} = \begin{matrix} \boldsymbol{b} \\ \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \end{matrix} \quad (\textit{inconsistent})$$

A already has orthogonal columns, so it's easy to find

$$\widehat{\boldsymbol{b}} = \frac{\boldsymbol{b}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} + \frac{\boldsymbol{b}\cdot\boldsymbol{v_2}}{\boldsymbol{v_2}\cdot\boldsymbol{v_2}}\boldsymbol{v_2} = \frac{0}{2}\begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} + \left(-\frac{3}{13}\right)\begin{bmatrix} 0 \\ 2 \\ 3 \\ 0 \end{bmatrix}$$

so $\widehat{\boldsymbol{b}} = \begin{bmatrix} 0 \\ -\frac{6}{13} \\ -\frac{9}{13} \\ 0 \end{bmatrix}$. Solving $A\boldsymbol{x} = \widehat{\boldsymbol{b}}$ is easy because

we can see from the equation for $\widehat{\boldsymbol{b}}$ that

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{3}{13} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{6}{13} \\ -\frac{9}{13} \\ 0 \end{bmatrix}$$

$$\| \qquad \begin{bmatrix} \frac{\boldsymbol{b}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1} \\ \frac{\boldsymbol{b}\cdot\boldsymbol{v}_2}{\boldsymbol{v}_2\cdot\boldsymbol{v}_2} \end{bmatrix}$$

For the equation $A\boldsymbol{x} = \boldsymbol{b}$

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

the least squares error (in using $\widehat{\boldsymbol{x}}$ as a "solution")

$$= \| A\,\widehat{\boldsymbol{x}} - \boldsymbol{b} \| = \| \widehat{\boldsymbol{b}} - \boldsymbol{b} \|$$

$$= \| \begin{bmatrix} 0 \\ -\frac{6}{13} \\ -\frac{9}{13} \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \|$$

$$= \sqrt{(0-1)^2 + (-\tfrac{6}{13} - 0)^2 + (-\tfrac{9}{13} + 1)^2 + (0-1)^2}$$

$$\approx 1.5191$$

In this example, there is <u>only one</u> $\widehat{\boldsymbol{x}}$; so $\boldsymbol{x}$ is <u>not</u> a least squares solution if $\boldsymbol{x} \neq \widehat{\boldsymbol{x}}$, That means that $\boldsymbol{x}$ does worse as an "approximate solution" to $A\boldsymbol{x} = \boldsymbol{b}$:

least squares error = $\| A\,\widehat{\boldsymbol{x}} - \boldsymbol{b} \| = 1.5191 \quad < \|A\boldsymbol{x} - \boldsymbol{b}\|$

(Discussed in previous lecture)

Easier way to find least squares solution for $A\boldsymbol{x} = \boldsymbol{b}$

$A\boldsymbol{x} = \widehat{\boldsymbol{b}}$ has same solutions as $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$

↑ normal equation for $A\boldsymbol{x} = \boldsymbol{b}$

(*no need to find* $\widehat{\boldsymbol{b}}$ *first, so no need to find an orthogonal hasis for Col* A):

Example Find least squares solutions for $A\boldsymbol{x} = \boldsymbol{b}$ (*inconsistent*)

$$\begin{bmatrix} -5 & 2 & -3 & 2 & -1 \\ 1 & 1 & 2 & 1 & 0 \\ 3 & 1 & 4 & 1 & 1 \\ -1 & 1 & 0 & 1 & 0 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0 \\ 3 \\ 6 \\ 2 \end{bmatrix}$$

Solve normal equation $A^TA\,\boldsymbol{x} = A^T\boldsymbol{b}$

$$\begin{bmatrix} 36 & -7 & 29 & -7 & 8 \\ -7 & 7 & 0 & 7 & -1 \\ 29 & 0 & 29 & 0 & 7 \\ -7 & 7 & 0 & 7 & -1 \\ 8 & -1 & 7 & -1 & 2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 19 \\ 11 \\ 30 \\ 11 \\ 6 \end{bmatrix}$$

Row reducing the augmented matrix gives

$$\begin{bmatrix} \mathbf{1} & 0 & 1 & 0 & 0 & \frac{19}{34} \\ 0 & \mathbf{1} & 1 & 1 & 0 & \frac{41}{17} \\ 0 & 0 & 0 & 0 & \mathbf{1} & \frac{67}{34} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$ There are 2 free variables: $\boldsymbol{x_3}$, $\boldsymbol{x_4}$

General solution is $\boldsymbol{x} = \widehat{\boldsymbol{x}}$, where

$$\widehat{\boldsymbol{x}} = \begin{bmatrix} \frac{19}{34} \\ \frac{41}{17} \\ 0 \\ 0 \\ \frac{67}{34} \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

So there are infinitely many "least squares solutions"

Any two of them, say $\widehat{\boldsymbol{x}_1}$ and $\widehat{\boldsymbol{x}_2}$, produce the same (smallest) least squares error:

$$|| A\widehat{\boldsymbol{x}_1} - \boldsymbol{b} || = || \widehat{\boldsymbol{b}} - \boldsymbol{b} || = || A\widehat{\boldsymbol{x}_2} - \boldsymbol{b} ||$$

<u>If there are infinitely many least squares solutions, then they all produce the same least squares error</u>.

Example Find the least squares solution for

$$\begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Write the normal equations:

$$A^T A = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 13 & -1 \\ -1 & 2 \end{bmatrix}$$

and

$$A^T \boldsymbol{b} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

so the normal equations are

$$\begin{bmatrix} 13 & -1 \\ -1 & 2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

$A^T A$ is invertible

$$\begin{bmatrix} 13 & -1 \\ -1 & 2 \end{bmatrix}^{-1} = \tfrac{1}{25} \begin{bmatrix} 2 & 1 \\ 1 & 13 \end{bmatrix},$$

so there is a unique least squares solution:

$$\widehat{\boldsymbol{x}} = \tfrac{1}{25} \begin{bmatrix} 2 & 1 \\ 1 & 13 \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} \\ \frac{1}{5} \end{bmatrix}$$

Least squares error is

(*) $$\| \boldsymbol{b} - A\widehat{\boldsymbol{x}} \| = \| \boldsymbol{b} - \widehat{\boldsymbol{b}} \|$$

$$= \| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} \frac{2}{5} \\ \frac{1}{5} \end{bmatrix} \| = \| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \| = 0$$

Least squares error $= 0$ means that

$$\boldsymbol{b} = \widehat{\boldsymbol{b}}, \qquad \text{so } A\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{b}} = \boldsymbol{b}:$$

the least squares solution $\widehat{\boldsymbol{x}} = \begin{bmatrix} \frac{2}{5} \\ \frac{1}{5} \end{bmatrix}$ is really an exact

solution (which we could have noticed in the first place!).

Simple Linear Regression Model (discussed in class)

i	x_i	y_i
1	36	6
2	82	14
3	45	5
4	49	13
5	21	5
6	58	14
7	73	11
8	85	18
9	52	6
10	24	8

find straight line $\beta_0 + \beta_1 x$ that "best fits" the data

Here, the data is from an experiment about foods and hyperactivity: x_i represents the number of grams of food containing certain preservatives eaten by a child and y_i is a "hyperactivity" score for the same child.

If there were a solution $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ for the system

$$\begin{bmatrix} 1 & 36 \\ 1 & 82 \\ 1 & 45 \\ 1 & 49 \\ 1 & 21 \\ 1 & 58 \\ 1 & 73 \\ 1 & 85 \\ 1 & 52 \\ 1 & 24 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 6 \\ 14 \\ 5 \\ 13 \\ 5 \\ 14 \\ 11 \\ 18 \\ 6 \\ 8 \end{bmatrix},$$

X (model matrix) $\quad \boldsymbol{y}$ (observation vector)

then the data points would all lie exactly on the line $\beta_0 + \beta_1 x$ (*see class notes*)

For this data, that's false – the system is inconsistent. The "best we can do" is to find a least squares solution $\widehat{\boldsymbol{\beta}}$

The normal equation $X^T X\boldsymbol{\beta} = X^T \boldsymbol{y}$ **is**

$$\begin{bmatrix} 10 & 525 \\ 525 & 32085 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 100 \\ 5980 \end{bmatrix}$$

$X^T X$ is invertible, so there's a unique least squares solution:

$$\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} = \begin{bmatrix} 10 & 525 \\ 525 & 32085 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 5980 \end{bmatrix}$$

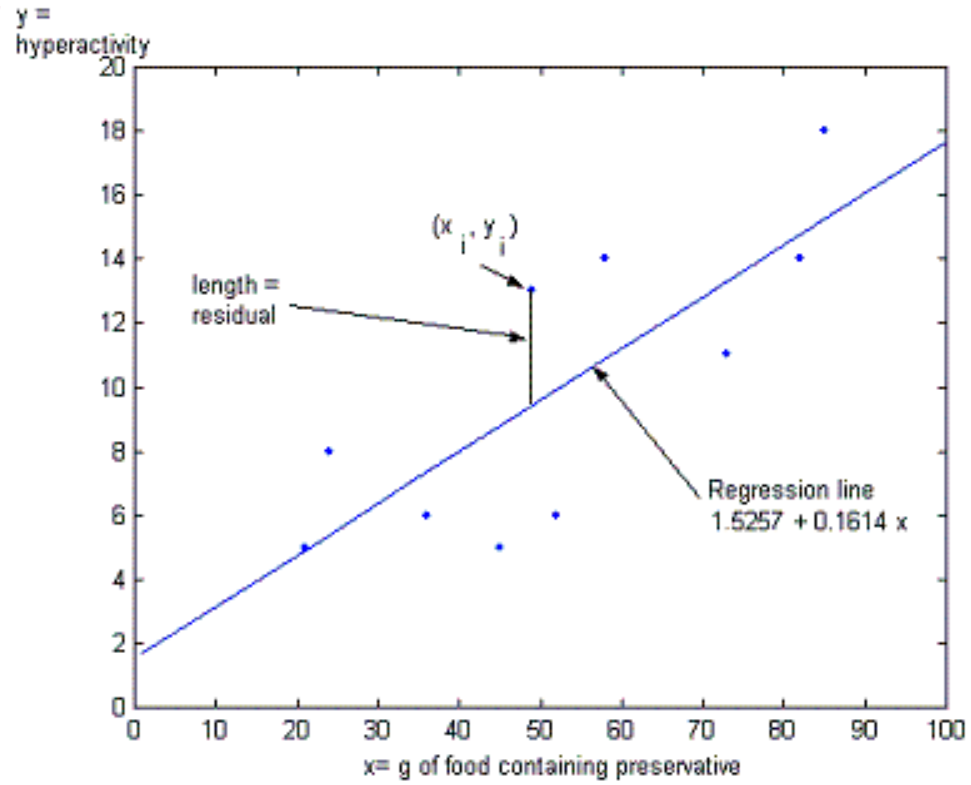
$$= \tfrac{1}{45225} \begin{bmatrix} 32085 & -525 \\ -525 & 10 \end{bmatrix} \begin{bmatrix} 100 \\ 5980 \end{bmatrix} \approx \begin{bmatrix} 1.5257 \\ 0.1614 \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}$$

(*rounded to 4 decimal places*)

The regression line is

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x \approx 1.5257 + 0.1614x$$

The data points and the regression line are pictured below.



Each $y_i - (\beta_0 + \beta_1 x_i)$ is called a residual (see picture above). For each x_i, the residual is the difference "observed-predicted" $=$ (observed data value y_i) $-$ (y value predicted for x_i using the line $y = \beta_0 + \beta_1 x$ to make the prediction).

What does it mean that the regression line is the "best fitting" line?

Using the coefficients from $\boldsymbol{\beta} =$ least squares solution $\widehat{\boldsymbol{\beta}}$ minimizes $||\,\boldsymbol{y} - X\boldsymbol{\beta}||$

If we think of an "error vector" $\boldsymbol{\epsilon} = \boldsymbol{y} - X\boldsymbol{\beta}$ (so $X\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{y}$).
then the least square solution $\widehat{\boldsymbol{\beta}}$ minimizes $||\,\boldsymbol{\epsilon}\,|| =$ the size of the error vector

$$= \left|\left|\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{bmatrix}\right|\right| = \left|\left|\begin{bmatrix} y_1 - (\beta_0 + \beta_1 x_1) \\ \vdots \\ y_i - (\beta_0 + \beta_1 x_i) \\ \vdots \\ y_n - (\beta_0 + \beta_1 x_n) \end{bmatrix}\right|\right| = \sqrt{\sum_{ui=1}^{n} (y_i - (\beta_0 + \beta_1 x_i)^2)}$$

In other words, the regression line is chosen in the way that minimizes the sum of the squared residuals.

Of course, this happens when $\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i)^2)$ is

minimized. Loosely speaking, the regression line minimizes a sum of squares so it may also be called the "least squares line"